13

CLAIMS:

1.   A method of determining the likelihood of a candidate composition comprising:

selecting a reference group of known characterised compositions;

establishing statistical characteristics relating to components of or other features

5 of the known characterised composition;

comparing the statistical characteristics of the known characterised compositions with corresponding components or features in the candidate compositions to establish a likelihood of those compositions occurring.

10 2.   A method as claimed in claim 1 wherein the candidate composition is a glycan and the reference group of known characterised compositions is a reference group of glycans, the method comprising

providing a search mass of a glycan whose composition is to be determined;

generating a list of possible glycan compositions made up of components,

15 including monosaccharides, whose total mass is within a predetermined tolerance of the search mass;

selecting a reference group of known characterised glycan compositions of approximately similar mass to the search mass;

establishing the mean and standard deviation of each component appearing in

20 the reference group of the known characterised glycan compositions;

for each candidate glycan composition calculating a partial score for each component in that theoretical glycan candidate, the partial score being calculated from the mean and standard deviation of the component appearing in the reference group and which provides a measure of the likelihood of that component being present in the

25 candidate composition ;

combining the partial scores to provide an indication of the likelihood of that candidate glycan composition occurring.

3.   A method as claimed in claim 2 wherein the partial scores for each component

30 are based on the difference between the observed number of the component in the candidate glycan composition and the mean for that component in the reference group, divided by the standard deviation and wherein the combining of the partial scores is carried out by multiplying the partial scores together.

35 4.   A method as claimed in claim 2 wherein the partial score for each component is calculated according to the equation:-

$$partialscore_{monosac} = \frac{\left|mean_{monosac} - observed_{monosac}\right|}{stdev_{monosac}}$$

where $mean_{monosac}$ is the mean number of the given monosaccharide in the reference data set; $mean_{monosac}$ is the number of the given monosaccharide in the candidate glycan composition; and $stddev_{monosac}$ is the standard deviation of the given monosaccharide in the reference data set.

5.   A method as claimed in claim 2 wherein the partial score for each component is calculated according to the equation:-

$$PartialScore_m = \frac{e^{-\frac{1}{2}(StDevScore_m)^2}}{\sqrt{2\pi} \times stdev_m}$$

where $StDevScore_m = Abs(count_m - mean_m)/stdev_m$

6.   A method as claimed in claim 5 wherein the biological index is calculated according to the equation:

$$BiologicalIndex = \frac{1}{\ln\left(\prod_{m \in monosaccharides} PartialScore_m\right)}$$

7.   A method as claimed in any one of claims 2 to 6 wherein the predetermined tolerance of the search mass is within +/- 400Da, preferably +/- 200Da.

8.   A system for determining the likelihood of a candidate composition comprising a computer means running software implementing the method of any one of claims 1 to 6.

9.   A method of determining the likelihood of a candidate glycan composition using a system as claimed in claim 8 when dependent on any one of claims 2 to 6 including :-

        inputting a search mass;

        inputting a search mass tolerance;

        inputting a biological index cut off; and

        inputting a maximum value for each component in the candidate composition.